MAKING DIFFICULT MODEL COMPARISONS GRAPHICALLY

David F. Findley
Bureau of the Census, Washington, D.C. 20233

1. INTRODUCTION

Suppose one wishes to decide between two model families, not necessarily nested and not necessarily correct, for observed data $y_1,...,y_N$. Conceptually, there are two possible situations: either the theoretically best-fitting models from the competing classes fit (i) equally well, or (ii) one model class is capable of providing a better fit than the other. In case (i), variability arising from parameter estimation can still cause one of the two classes to be preferred, and this is the playing field of the null hypothesis in classical statistical tests. However, general statistical procedures for identifying the preferred class in this case seem to require rather strong assumptions about the approximate correctness of the models. By contrast, as the results of this paper show, there are theoretically founded procedures with relatively weak requirements for making the more fundamental decisions, does (i) hold, or (ii)? -- and, if (ii), which model class provides the better fit?

As section 2 explains, these two questions are answered by deciding if the log-likelihood difference $\hat{L}_N^{(1,2)}$ of the maximum likelihood values from the two families diverge to an infinite value as $N \rightarrow \infty$ and, if so, whether to $+\infty$ or $-\infty$. The obvious diagnostic for such divergence is a graph of the log-likelihood differences from an appropriately selected increasing sequence of subsets of the observed data set. However, the calculation of the sequence of m.l.e.'s and the likelihood values required for such a graph can be quite demanding computationally and therefore poorly suited to interactive modeling. In section 3, we will describe a related graphical diagnostic which is suited to interactive modeling and is especially convenient because it requires only quantities which are usually available from the full-data-set likelihood maximizations. Also, a condition, (3.5), is given which guarantees that this diagnostic describes the relevant behavior of $\hat{L}_N^{(1,2)}$ for large enough sample sizes N. The proposition of section 4 shows that (3.5) can be verified under rather weak assumptions: for example, it is not required that the maximum likelihood parameter estimates converge uniquely. Section 5 discusses some subtle points concerning log-likelihood ratios for nonstationary time series models. Section 6 describes the results of applying the graphical diagnostics to the comparison of ARIMA models with structural component models for 10 U.S. Census Bureau time series.

2. WHEN AND HOW DO LOG-LIKELIHOOD-RATIOS DIVERGE TO $\pm\infty$?

Let $L_N[\theta^{(1)}]$, $\theta^{(1)} \in \Theta^{(1)}$ and $L_N[\theta^{(2)}]$, $\theta^{(2)} \in \Theta^{(2)}$ be two parametric families of log-likelihoods defined by competing models for observed random variates $y_1,...,y_N$. We do not assume that the competing log-likelihood functions have a similar form or are related in any way. A notation that more strongly emphasized possible differences of form would be $L^{(j)}[\theta^{(j)}]$, j=1,2, but we will avoid the duplicated superscript to reduce notational complexity, anticipating that this will not cause confusion for the reader. Usually each family of log-likelihood functions has an associated family of non-random entropy functions $E_\infty[\theta^{(j)}]$, $\theta^{(j)} \in \Theta^{(j)}$ which are the limits (existing with probability one) of the sample-size-normalized log-likelihood functions,

$$\text{(2.1)} \qquad E_\infty[\theta^{(j)}] = \lim_{N\to\infty} N^{-1}L_N[\theta^{(j)}] \ (j = 1,2) \ \text{(w.p.1)}.$$

If maximum likelihood estimates $\hat{\theta}_N^{(j)}$ exist and if we define

$$\text{(2.2)} \qquad E_\infty^{(j)} \equiv \sup_{\theta^{(j)}\in\Theta^{(j)}} E_\infty[\theta^{(j)}],$$

then, ordinarily (with p-lim denoting convergence in probability)

$$\text{(2.3)} \qquad \text{p-lim}_{N\to\infty} N^{-1}L_N[\hat{\theta}_N^{(j)}] = E_\infty^{(j)}$$

will hold for j = 1,2: see White (1990) and the proof of Theorem 7.4.10 of Hannan and Deistler (1988) where (2.3) is established for invertible ARMA models without the assumption that the model classes under consideration contain the true model. We will use the abbreviations $\hat{L}_N^{(j)} \equiv L_N[\hat{\theta}_N^{(j)}]$, i = 1,2, and, for the log-likelihood difference (= log likelihood-ratio),

$$\hat{L}_N^{(1,2)} \equiv \hat{L}_N^{(1)} - \hat{L}_N^{(2)}.$$

Then, from (2.3), we obtain

$$\text{(2.4)} \qquad \text{p-lim}_{N\to\infty} N^{-1}\hat{L}_N^{(1,2)} = E_\infty^{(1)} - E_\infty^{(2)},$$

and, as a consequence,

(2.5) <u>If $E_\infty^{(1)} \neq E_\infty^{(2)}$, then</u> $\text{p-lim}_{N\to\infty} \hat{L}_N^{(1,2)} = \pm\infty$,

<u>where the sign of the limit and its asymptotic slope are the sign and the slope of $(E_\infty^{(1)} - E_\infty^{(2)})N$. Thus, $\hat{L}_N^{(1,2)} = O_p(N)$.</u>

This result has a particularly straightforward interpretation when Gaussian likelihood functions are used (without assuming the data are Gaussian), because then, usually,

$$\text{(2.6)} \qquad E_\infty^{(j)} = -\log 2\pi e \sigma_\infty^{(j)},$$

where $\sigma_\infty^{(j)2}$ is the variance of the asymptotic "residuals" process associated with the best fitting model(s) in the class being considered: for example, for competing Gaussian ARMA (or ARIMA) time series models, $\sigma_\infty^{(j)2}$ is the variance of the one-step-ahead forecast error process for a model determined by a limiting value $\theta_\infty^{(j)}$ of $\hat{\theta}_N^{(j)}$, j=1,2. Under (2.6), the sign of $E_\infty^{(1)} - E_\infty^{(2)}$ is that of $\sigma_\infty^{(2)} - \sigma_\infty^{(1)}$. This means that the model class with better one-step-ahead forecasting properties for the observed series is the model class whose log-likelihood function will dominate the log-likelihood difference as $N \to \infty$.

3. GRAPHICAL DIAGNOSTICS FOR MODEL COMPARISON: DETECTING DIVERGENCE PROPERTIES OF THE LOG-LIKELIHOOD RATIO.

The result (2.5) immediately suggests a graphical method for detecting whether or not one of two log-likelihoods will, in large samples, strongly dominate the other and thereby identify itself as the model which is to be preferred: reestimate the model over an increasing sequence of subsets of the available observations and plot the resulting sequence of likelihood ratios as a function of sample size, looking for a linear trend up or down in the later part of the graph. In the case where the data index has a natural ordering, as with time series, this procedure would suggest calculating $\hat{\theta}_M^{(1)}$ and $\hat{\theta}_M^{(2)}$ from $y_1,...,y_M$ for, say, $N/2 < M \le N$, and plotting the log-likelihood differences

$$(3.1) \qquad \hat{L}_M^{(1,2)}, \; N/2 < M \le N$$

as a function of increasing M. However, the calculation of the quantities in (3.1) could be time consuming and also quite inconvenient to do with some software packages. We shall argue in this and the next section that plotting

$$(3.2) \qquad L_M[\hat{\theta}_N^{(1)}] - L_M[\hat{\theta}_N^{(2)}], \; N/2 < M \le N$$

versus M is much less burdensome computationally, yet also, when N is large enough, equally informative about the divergence properties of $\hat{L}_N^{(1,2)}$. Calculating (3.2) rather than (3.1) has the obvious advantage that only a single likelihood maximization is necessary for each model class to obtain $\hat{L}_N^{(1)}$ and $\hat{L}_N^{(2)}$. Less obvious is the fact that the quantities in (3.2) are often immediately available as a byproduct of the calculation of $\hat{L}_N^{(i)}$, $i = 1,2$. This is easy to see when the data are modeled as though they are independent and identically distributed: in this case the log-likelihoods have the form

$$L_N[\theta^{(j)}] = \Sigma_{n=1}^{N} \log g[\theta^{(j)}](y_n)$$

and, for any $M \le N$,

$$(3.3) \qquad L_M[\hat{\theta}_N^{(j)}] = \Sigma_{n=1}^{M} \log g[\hat{\theta}_N^{(j)}](y_n),$$

for $j = 1,2$.

In the case of dependent time series data modeled as a Gaussian ARMA or ARIMA model, there is an analogue of (3.3) which arises from the conditional decomposition

$$L_N[\theta^{(j)}] = \log g[\theta^{(j)}](y_1) + \Sigma_{n=2}^{N} \log g[\theta^{(j)}](y_n|y_{n-1},...,y_1)$$

and takes the form

$$(3.4) \qquad L_M[\hat{\theta}_N^{(j)}] = -\tfrac{1}{2} \Sigma_{n=1}^{M} \{ \log 2\pi\sigma_{n|n-1}^2[\hat{\theta}_N^{(j)}] + (y_n - y_{n|n-1}[\hat{\theta}_N^{(j)}])^2 / \sigma_{n|n-1}^2[\hat{\theta}_N^{(j)}] \}$$

In this expression, $y_{n|n-1}[\theta]$ denotes the linear function of $y_{n-1},...,y_1$ which would provide the best predictor of y_n if the data were Gaussian with the mean and covariance structure specified by θ (or, when n=1, the mean specified for y_1 by θ); $\sigma_{n|n-1}^2[\theta]$ is the function of θ given by the mean square of $y_n - y_{n|n-1}[\theta]$ calculated with respect to the joint density $\exp(L_N[\theta])$. All of these quantities can be calculated from one pass over the data with the Kalman filter algorithm, given a state space representation of the time series model and a suitable initialization, see Bell and Hillmer (1990), for example. If the Kalman filter has been used to evaluate the likelihood function in the maximization routine, all of the quantities required for (3.4) and (3.2) will be available after the last maximization step.

Some graphs of (3.1) and (3.2) for competing models (described in section 6) are presented on the last page. There is further discussion of these figures in section 6.

In the next section, we shall demonstrate the large-sample equivalence of the sets of statistics (3.1) and (3.2) by verifying the condition

$$(3.5) \qquad \lim_{M\to\infty} \sup_{N \ge M} |M^{-1} L_M[\hat{\theta}_N^{(j)}] - E_\infty^{(j)}| = 0 \quad \text{(w.p.1)}$$

for j=1,2. This condition implies that either set of statistics can be used to determine the sign of $E_\infty^{(1)} - E_\infty^{(2)}$ if this quantity is non-zero. (To justify the use of (3.2), it would suffice to establish a weaker version of (3.5) with N ranging only over $M \le N < 2M$.)

4. VERIFICATION OF (3.5): A GENERAL RESULT

One attractive feature of the result of this section is that it accommodates the situation, observed by Kabaila (1983) to occur with an incorrect first order moving average time series model, wherein the m.l.e.'s $\hat{\theta}_N$ do not converge to a single value, but rather have a set of limiting values,

$$(4.1) \qquad \Theta_0 \equiv \{\theta: E_\infty[\theta] = E_\infty\}.$$

For a set F containing Θ_0, we shall write

$$(4.2) \qquad \hat{\theta}_N \to \Theta_0 \text{ (in F, w.p.1)}$$

if, excluding realizations $\{y_n(\omega)\}_{1 \le n < \infty}$ of $\{y_n\}_{1 \le n < \infty}$ which form an event of probability 0, every subsequence of $\{\hat{\theta}_N(\omega)\}_{1 \le N < \infty}$ contains a subsequence which is in F and which converges to a point in Θ_0. This is equivalent to saying that, given any neighborhood V of Θ_0 in F, the probability is 1 that only finitely many of the events $\hat{\theta}_N \notin V$, N = 1,2,..., occur. The result we are after is the following.

Proposition 4.1. Suppose there is a set F containing the $E_\infty[\theta]$-minimizing set Θ_0 defined in (4.1) above such that (i) with probability one, the log-likelihood functions $L_N[\theta]$, $N \ge N_0$ are continuous on F; (ii) $N^{-1}L_N[\theta]$ converges uniformly to $E_\infty[\theta]$ on F (w.p.1); and (iii) the condition (4.2) is satisfied. Then (3.5) holds.

Proof. It follows from (i) and (ii) that $E_\infty[\theta]$ is continuous on F. Given $\delta > 0$, the set $V = \{\theta \in F: |E_\infty[\theta] - E_\infty| < \delta/2\}$ is thus an open set in F containing Θ_0. Therefore, by (4.2), the probability is one that only finitely many of the events $\theta_N \notin V$ occur and also, by (ii), that only finitely many of the events

$$\sup_{\theta \in V}|M^{-1}L_M[\theta] - E_\infty[\theta]| \geq \delta/2 \quad (M = 1,2,\ldots)$$

occur. Since

$$|M^{-1}L_M[\theta] - E_\infty| \leq |M^{-1}L_M[\theta] - E_\infty[\theta]| + |E_\infty[\theta] - E_\infty| ,$$

it follows that, with probability one, at most finitely many of the events

$$\sup_{N \geq M}|M^{-1}L_M[\theta_N] - E_\infty| \geq \delta \quad (M = 1,2,\ldots)$$

occur. This establishes the condition (3.5).

In many situations, the condition (ii) is a uniform law of large numbers, see Poetscher and Prucha (1989) and their references.

In Findley (1990b), Proposition 4.1 is utilized to verify (3.5) for invertible ARMA models.

5. LOG-LIKELIHOODS FOR ARIMA MODELS AND (3.5)

The models we wish to compare in section 6 are nonstationary ARIMA models. The nonstationarity introduces some subtle complications which we address in this section. Consider the situation in which a stationarizing backshift-operator polynomial $\delta(B)$ of degree d with $\delta(0) = 1$ is applied to the observed series $y_1,\ldots,y_N$ to obtain the data $w_n \equiv \delta(B)y_n$, $n = d+1,\ldots,N$ which are actually modeled, from a log-likelihood family $L_{N,d}[\theta] = L[\theta](w_{d+1},\ldots,w_N)$. If we let $L_d(y_1,\ldots,y_d)$ denote the (unknown) log-density of $y_1,\ldots,y_d$, then

$$L_N[\theta] \equiv L_{N,d}[\theta] + L_d$$

is a log-likelihood for $y_1,\ldots,y_N$, in the sense that

$$\int_{\mathbb{R}^N} \exp(L_N[\theta])dy_1\cdots dy_N = 1 \quad ,$$

if the integral of $\exp(L_{N,d}[\theta])$ over $\mathbb{R}^{N-d}$ is 1: indeed, since the Jacobian of the transformation

$$(y_1,\ldots,y_N) \rightarrow (y_1,\ldots,y_d,w_{d+1},\ldots,w_N)$$

is 1, we have

$$\int_{\mathbb{R}^N} \exp(L_N[\theta])dy_1\cdots dy_N =$$

$$\int_{\mathbb{R}^d} \exp(L_d)dy_1\cdots dy_d \int_{\mathbb{R}^{N-d}} \exp(L_{N,d}[\theta])dw_{d+1}\cdots dw_N$$

$$= \int_{\mathbb{R}^{N-d}} \exp(L_{N,d}[\theta])dw_{d+1}\cdots dw_N \ .$$

For a more general discussion, see Findley (1990a). Note that $L_N[\theta^{(0)}]$ will be the correct log density for $y_1,\ldots,y_N$ if $L_{N,d}[\theta^{(0)}]$ is the correct density of $w_{d+1},\ldots,w_N$, and if, at the same time, $y_1,\ldots,y_d$ are independent of the w_n process. This independence is usually assumed, in order to insure that the one-step-ahead forecast-error (innovations) process of y_n coincides with that of w_n, see Bell (1984).

Suppose we have candidate transformations $\delta^{(j)}(B)$ of degree $d^{(j)}$ and candidate m.l.e. models $\hat{L}^{(j)}_{N,d^{(j)}} \equiv L_{N,d^{(j)}}[\theta^{(j)}_{N,d^{(j)}}]$ for the data $w^{(j)}_n = \delta^{(j)}(B)y_n$, $n = d^{(j)}+1,\ldots,N$, $j = 1,2$. Then the log-likelihood difference $\hat{L}^{(1,2)}_N = L_N[\theta^{(1)}_{N,d^{(1)}}] - L_N[\theta^{(2)}_{N,d^{(2)}}]$ satisfies

$$(5.1) \qquad \hat{L}^{(1,2)}_N = \hat{L}^{(1)}_{N,d^{(1)}} - \hat{L}^{(2)}_{N,d^{(2)}} + \{L_{d^{(1)}} - L_{d^{(2)}}\},$$

so $\hat{L}^{(1,2)}_N$ is known to within a summand which is a function of $y_1,\ldots,y_{\max\{d^{(1)},\ d^{(2)}\}}$.

Of course, when $d^{(1)} = d^{(2)}$, then $\hat{L}^{(1,2)}_N$ is known,

$$(5.2) \qquad \hat{L}^{(1,2)}_N = \hat{L}^{(1)}_{N,d^{(1)}} - \hat{L}^{(2)}_{N,d^{(2)}} \quad (d^{(1)} = d^{(2)}),$$

and the graphical procedures of section 3 can be applied. In fact, since $L_{d^{(1)}} - L_{d^{(2)}}$ does not change with N, it follows from (5.2) that, whether or not $d^{(1)} = d^{(2)}$,

$$(5.3) \quad \hat{L}^{(1,2)}_N \rightarrow \pm\infty \text{ if and only if } \hat{L}^{(1)}_{N,d^{(1)}} - \hat{L}^{(2)}_{N,d^{(2)}} \rightarrow \pm\infty,$$

so the graph of $\hat{L}^{(1)}_{M,d^{(1)}} - \hat{L}^{(2)}_{M,d^{(2)}}$, $N/2 < M \leq N$ can be examined to see if an ultimate direction for $\hat{L}^{(1,2)}_N$ is suggested.

When $d^{(1)} \neq d^{(2)}$, we shall refer to the quantities in (5.4) and (5.5) below as pseudo-log-likelihood-ratios (pseudo-LLR's):

$$(5.4) \qquad \hat{L}^{(1)}_{M,d^{(1)}} - \hat{L}^{(2)}_{M,d^{(2)}}, \quad N/2 < M \leq N$$

and

$$(5.5) \qquad L_{M,d^{(1)}}[\theta^{(1)}_N] - L_{M,d^{(2)}}[\theta^{(2)}_N], \quad N/2 < M \leq N.$$

In Findley (1990b), graphs of (5.4) and (5.5) are given for two series with models having $d^{(1)} \neq d^{(2)}$, see Figs. 3c, d and 7c, d.

Remark 5.1. If the approach described here to defining $\hat{L}^{(1,2)}_N$ for ARIMA models is used, there are some implications concerning the applicability of Akaike's AIC criterion. Consider the difference of AIC values,

$$(5.6) \qquad \Delta AIC^{(1,2)}_N \equiv -2\hat{L}^{(1,2)}_N + 2(\dim\theta^{(1)} - \dim\theta^{(2)}),$$

where $\dim\theta^{(j)}$ denotes the number of estimated parameters in the j-th model family, j=1,2. It is clear from (5.1) than when $d^{(1)} \neq d^{(2)}$, since $L_{d^{(1)}} - L_{d^{(2)}}$ has non-zero mean, the calculable analogue of (5.6),

$$(5.7) \qquad -2\{\hat{L}^{(1)}_{N,d^{(1)}} - \hat{L}^{(2)}_{N,d^{(2)}}\} + 2(\dim\theta^{(1)} - \dim\theta^{(2)}) ,$$

will not have the same asymptotic mean as the uncalculable quantity $\Delta AIC_N^{(1,2)}$ when the means of the sequence $\hat{L}^{(1,2)}_N$ are bounded. As a consequence, the bias calculations motivating the use of AIC (see Findley (1985) and Findley and Wei (1989)) do not support the use of (5.7). Of course, if $\mathcal{L}_N^{(1,2)} \to \pm\infty$, the finite bias correction term $2(\dim\theta^{(1)} - \dim\theta^{(2)})$ is inconsequential.

6. COMPARISONS OF ARIMA AND COMPONENT MODELS

Bell and Pugh (1989) compared ARIMA models fit individually to a large set of log-transformed economic time series with the best-fitting (that is, maximum likelihood) basic structural component model (BSM) of Harvey and Todd (1983) for each series. This model can be written, for our purpose, as

$$y_n = S_n + T_n + I_n \tag{6.1}$$

where S_n, T_n and I_n are independent series presumed to satisfy

$$(1+B + \ldots B^{11})S_n = e_{1n},\ e_{1n} \sim \text{i.i.d } N(0, \mu\sigma^2)$$

$$(1-B)^2T_n = (1-\eta)e_{2t},\ e_{2t} \sim \text{i.i.d } N(0, \gamma\sigma^2)$$

$$I_n \sim \text{i.i.d. } (0, \sigma^2)$$

If the estimated value of η exceeded 0.9, we often used a different model for T_n,

$$(1-B)T_n = C + e_{2n},\ e_{2n} \sim \text{i.i.d. } (0, \sigma^2)$$

with C a constant term, in order to avoid the technical problems which were discussed in section 5. We refer to this model as a modified component model.

In addition to the three components of (6.1), the models considered for most of the series have a mean component consisting of a sum of indicator variables for highly significant additive outliers, together with linear regression expressions modeling calendar effects, see Bell and Hillmer (1983). The theoretical discussion in the preceding sections concerned mean zero time series, so we need to say something about the additional assumptions and developments required to cover the situation of estimated mean functions. The estimation of the coefficients of the indicator variables has an asymptotically negligible effect on the likelihood function because these are localized to single observations. So, for theoretical purposes, we assume that such coefficients are fixed and not reestimated as $N \to \infty$. The calendar effect variables can be regarded as periodic with long periods, and they satisfy Grenander's conditions as discussed in Hannan (1973). Our method of simultaneously estimating regression and ARMA coefficients is described in Findley, Monsell, Otto, Bell and Pugh (1988). We shall assume that with properly chosen coefficients (perhaps zero) these variables completely describe the mean function of y_n, even though the remainder of the model might not completely describe the covariance structure of the series. With this assumption, the methods used for the proof of Theorem 4 of Hannan (1973) can be utlized to obtain a generalization of Proposition 4.1 which covers models with such mean functions.

Bell and Pugh used Akaike's AIC as the basic comparison statistic. Hence they used the sign of

$$\Delta AIC_N^{(1,2)} = -2\mathcal{L}_N^{(1,2)} + 2(\dim\theta^{(1)} - \dim\theta^{(2)})$$

to indicate the preferred model, the first model being preferred over the second if $\Delta AIC_N^{(1,2)} < 0$. Here $\dim\theta^{(j)}$ denotes the number of independent parameters in the j-th family. In our comparisons, the ARIMA model family is designated the first family (j=1) and the component or modified component model family is always the second family (j=2).

The values of $\Delta AIC_N^{(1,2)}$ and the interpretation of the graph of (3.2) are given in Table 1 for the 10 series from Bell and Pugh (1989) for which both the ARIMA and the (perhaps modified) structural model were invertible. For each series, the same transformation to stationarity was used for both models, $\delta^{(1)}(B) = \delta^{(2)}(B)$. The graphical diagnostic was interpreted as inconclusive (I) unless there was a possibly oscillatory but never-the-less clearly perceptible linear trend (upward, or downward) with non-zero slope over an interval of time which was large relative to any earlier time intervals in (N/2, N] during which the general movement was in the opposite direction.

On the last page, graphs of (3.1) and (3.2) are given for a series (cneths) for which these diagnostics favor the ARIMA model and also for a series (bdptrs) where no linear movement toward $+\infty$ or $-\infty$ is visible, with the result the graphical diagnostics are inconclusive. In the latter case, the fact that the graphs are usually well above the horizontal line at level $\dim\theta^{(1)} - \dim\theta^{(2)} = 1$ most of the time shows that AIC's preference for the ARIMA model is rather stable. The inconclusive situation is the natural one in which to apply AIC, see Findley and Wei (1989). Findley (1990b) describes the results of analyzing (3.1) and (3.2) for 40 series in the study by Bell and Pugh (1989) and presents a generalization to time series models of the test statistic of Vuong (1989) which offers an independent confirmation of the conclusions reached by (3.1) and (3.2). The diagnostics favor an ARIMA for 18 of the series, a component model for one of the series, and are inconclusive for 21 series, often because the graphs of (3.1) and (3.2) become level in the last few years of the series (1980-2).

7. DISCLAIMER

This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributed to the author and do not necessarily reflect those of the Census Bureau.

REFERENCES

Bell, W. R. (1984). Signal extraction for non-stationary time series. Annals of Statistics 12, 646-664.

Bell, W. R. and Pugh, M. G. (1989). Alternative approaches to the analysis of time series components. Proceedings of the Statistics Canada Symposium on

Analysis of Data in Time. Ottawa: Statistics Canada (to appear).

Bell, W. R. and Hillmer, S. C. (1983). Modeling time series with calendar variation. Journal of the American Statistical Association 78, 526-534.

Findley, D. F. (1990a). Conditional densities as densities, and likelihoods defined via singular transformations. (manuscript in preparation)

Findley, D. F. (1990b). Making difficult model comparisons. Statistical Research Division Research Report No. RR-90/11. Washington, D.C.: U.S. Bureau of the Census.

Findley, D. F. (1985) On the unbiasedness property of AIC for exact or approximating linear time series models. J. Time Series Anal. 6, 229-252.

Findley, D. F., Monsell, B. C., Otto, M. C., Bell, W. R. and Pugh, M. G. (1988) Toward X-12-ARIMA. Proceedings of the Fourth Annual Research Conference, 591-622. Washington, D.C.: Bureau of the Census.

Findley, D. F. and Wei, C. Z. (1989). Beyond chi-square: likelihood ratio procedures for comparing non-nested, possibly incorrect regressors. Statistical Research Division Research Report No. RR-89/08. Washington, D.C.: U.S. Bureau of the Census.

Hannan, E. J. (1973). The asymptotic theory of linear time series models. Journal of Applied Probability 10, 130-145.

Hannan, E. J. and Deistler, M. (1988). The Statistical Theory of Linear Systems. New York: Wiley.

Harvey, A. C. and Todd, P.H.J. (1983). Forecasting economic time series with structural and Box-Jenkins models: a case study (with discussion). Journal of Business and Economic Statistics 1, 299-315.

Kabaila, P. (1983). Parameter values of ARMA models minimizing the one-step-ahead prediction error when the true system is not in the model set. J. Appl. Prob. 20, 405-408.

Poetscher, B. M. and Prucha, I. R. (1989). A uniform law of large numbers for dependent and heterogenous data sets. Econometrica 57, 675-683.

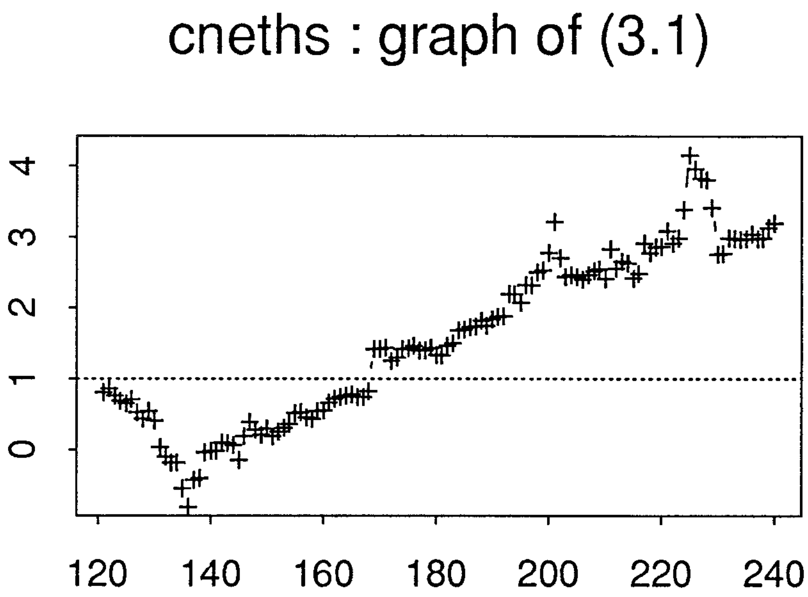
Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307-333.

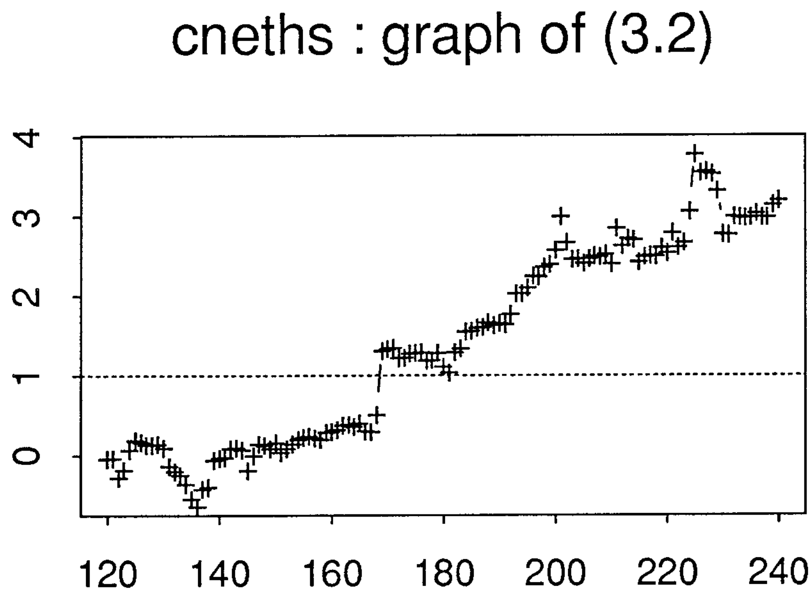
Table 1. ARIMA vs. Component Models

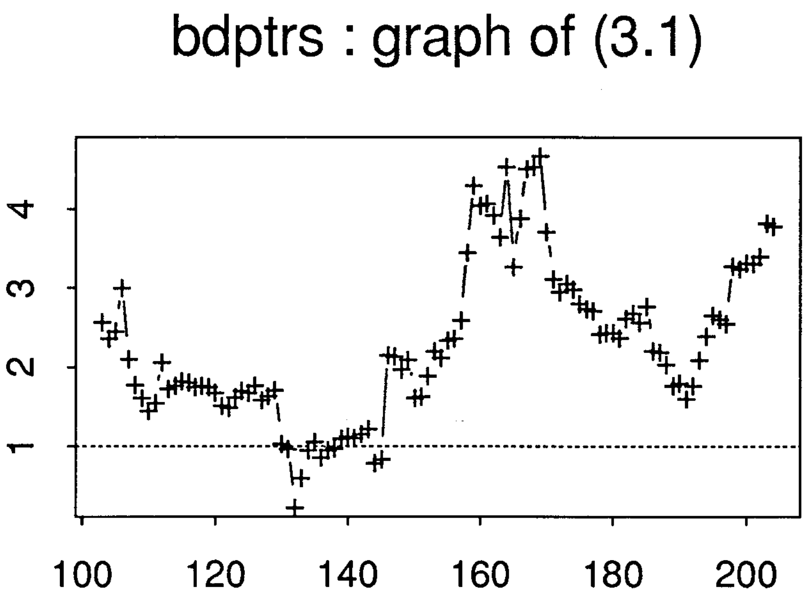
	$\Delta AIC_N^{(1,2)\dagger}$	GRAPH††
bdptrs†††	-8.9	I
bfrnrs†††	-2.2	I
bgasrs	-2.7	I
bmncrs†††	-22.1	A
cnctbp†††	-21.2	A
cncths†††	-27.7	A
cneths†††	-21.6	A
icmeti	-7.2	A
ifmeti	-29.0	A
itvrti	-20.6	A

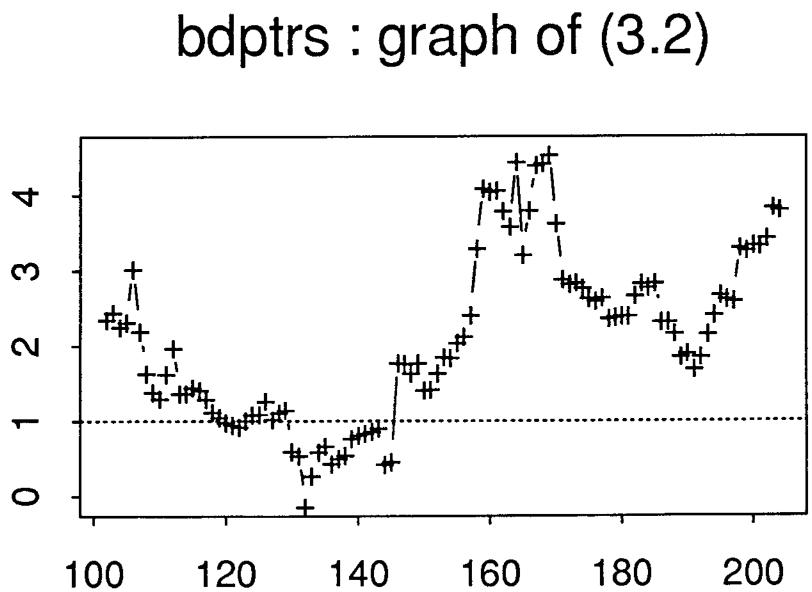
†: Negative values favor the ARIMA model
††: I = inconclusive; A = ARIMA model favored; C = component model favored.
†††: For these series, the modified component model was used.

bdptrs: Retail sales of department stores
bfrnrs: Retail sales of furniture stores
bgasrs: Retail sales of gasoline stations
bmncrs: Retail sales of men's and boys' clothing stores
cnctbp: Total North Central building permits
cncths: Total North Central housing starts
cneths: Total Northeast housing starts
icmeti: Total inventories of communications
ifmeti: Total inventories of farm machinery and equipment
itvrti: Total television and radio inventories


cneths : graph of (3.1)
0 1 2 3 4
120 140 160 180 200 220 240


cneths : graph of (3.2)
0 1 2 3 4
120 140 160 180 200 220 240


bdptrs : graph of (3.1)
1 2 3 4
100 120 140 160 180 200


bdptrs : graph of (3.2)
0 1 2 3 4
100 120 140 160 180 200